

Figure 1. The walls and ceiling of the Electroland Target breezeway are lined with an electronic “intelligent skin” composed of more than 24,000 LEDs behind translucent white glass, creating intense RGB light effects. Electroland software, integrated with the Tyzx technology, can modulate patterns depending on the number of visitors, the time of day and the level of activity in the room.



3-D Vision Gets Real

Next-generation technology sees and interacts in three dimensions.

by Ron Buck, Tyzx Inc.

Many classes of commercial and consumer products would benefit from the ability to visually sense their environment and react to what is happening in real time. These products would have to operate in dynamic, variably lit real-world environments. Robust sensing would require accurate visual analysis, which often entails three-dimensional data and high frame rates with low latency. In addition, the sensing system would have to be compact and low-cost, with low power demands. Although these requirements seem challenging, new technology is enabling affordable embedded products that see and interact with the world in three dimensions.

One application of visual sensing is a distributed person-tracking system that uses networks of smart sensors mounted on walls and ceilings. Electroland of Los Angeles — a creator of comprehensive and multidis-

ciplinary urban projects and scenarios — has produced for Target Corp. an interactive installation in the new observation deck of Rockefeller Center in New York (Figure 1).

People entering the space are assigned a “personality” and an individualized light pattern follows them around the space. The participants can influence and change the illuminated patterns in the room simply by moving around and by making various gestures.

The exhibit uses a 3-D vision system that can track up to thirty visitors with high accuracy and a very low error rate. This allows for precision down to 4 cm in translating individual motions into an engaging interactive experience. Person-tracking with a 3-D stereo camera can provide the location and movement of each individual precisely, even in a crowded environment. Although this particular application is primarily for entertainment, the same

technology also is being tested for homeland security and other surveillance and defense-related applications.

Benefits of 3-D tracking

Many vision-based tracking systems have traditionally relied only on 2-D data, which is inherently sensitive to changing lighting. In contrast, 3-D evidence can be derived from nonparametric stereo correlation, which is essentially a comparison of left and right images using pixel relative intensity and texture information. The result is that lighting and apparent color can be changing constantly without degrading tracking results.

The DeepSea G2 stereo vision system from Tyzx Inc. of Menlo Park, Calif., employs low-cost custom CMOS chips and sophisticated 3-D cameras (Figure 2). The 3-D data is much easier to segment into discrete subjects, which means that people

can be tracked even when they are substantially occluded from the view of a camera.

With a 2-D approach, the subject must be in the view of two or more cameras that also must remain well-registered to each other across a space. With 3-D vision, the subject needs to be only partially visible to a single stereo camera to be tracked; multiple cameras are used primarily to extend the operational footprint of the system and to contend with heavy occlusion.

In this system, a single 3-D stereo camera compares an image pair using stereo correlation. The resulting range image is a fine-grained representation of the distance to, and size of, each pixel in the scene. In contrast, multicamera tracking systems use discrete image features identified by conventional 2-D methods in monocular cameras, with subsequent triangulation. The 3-D cameras do not have to be well-registered because the imager's extrinsic parameters are fixed when the cameras are built.

The 3-D data produced by the stereo camera provides accurate information about the subject's size and precise location with respect to a world reference frame. In operation, the 3-D range and intensity images are first used in a background model to filter the incoming data and to identify pixels that are in motion, that is, those that do not match to the established background.

The resulting data, expressed as a binary foreground mask, is passed to ProjectionSpace, a postprocessing step that consolidates the massive amount of incoming 3-D data and presents it to the end application in a more directly usable fashion.

In this step, the orientation and position of the camera computed during registration are used to transform each 3-D foreground pixel coordinate into a coordinate system aligned with the floor and with other cameras. Objects with a high surface area perpendicular to the floor stand out, and people are especially easy to track because they are easily distinguished from the floor and other objects. In a 3-D person-tracking network, the segmented and



Figure 2. Cameras range from narrow (3-cm) baselines that can be used to study items only 15 to 20 cm from the camera to very long-baseline (up to 33-cm) specialty cameras that can be used to look at objects hundreds of yards away. The baselines on the cameras shown are 6 cm and 22 cm.

tracked results from many cameras are sent to a single PC system to produce a final tracking map, which can be used to clearly identify and follow all the individuals in a particular room or other given space.

Stereo vision

The system employs the company's application-specific integrated circuit for stereo correlation, a field programmable gate array, a digital signal processor and an embedded CPU running Linux. The 3-D person-tracking process described above is computationally intensive, but implementing the major steps — image rectification (removal of imaging imperfections), stereo correlation, background modeling and ProjectionSpace — as hardware visual primitives eases the computing requirements.

Stereo correlation collects the input from two imagers and integrates it into a single range image; background modeling takes the range and intensity data from the image to differentiate the person being tracked from his or her environment; and the postprocessor consolidates the incoming data so that it can be more easily handled and interpreted by a small, low-cost and low-power system. By accelerating these visual primitives in the cus-

tom application-specific integrated circuit and field programmable gate array, the CPU and digital signal processor are freed to host the remaining tracking software and other user applications.

Real-time tracking of people or other objects in motion requires 3-D data at high frame rates, and the large size of the 3-D data set, generated in real-time, can overwhelm large CPUs, let alone an embedded one. To mitigate this situation, the ProjectionSpace application programming interface normalizes and reduces the amount of 3-D data that the CPU must deal with.

First, rigid transform is applied to the 3-D data set from the stereo camera coordinate system to a user-specified coordinate system; for example, one aligned with the floor and in common with a world coordinate system shared by other stereo cameras. Second, the 3-D data set is quantized into 2- or 3-D cells, and the occupancy evidence in the cell is tested against a threshold — all of which is user-specified (Figure 3).

Embedding these visual primitives into custom hardware reduces the CPU workload and improves the tracking system's performance. This technology can track at twice the frame rate (30 fps) and with a fraction of the power as compared with

the company's first-generation system, which had only accelerated stereo correlation and a high-perfor-

mance Pentium processor.

This approach to computer vision solves technical challenges that made

3-D vision solutions unaffordable for volume commercial markets: performance, cost, size and power consumption. And the system remains flexible, allowing developers to adapt a wide variety of applications to the new platform. Further improvements in performance, power and cost may come as new visual primitives are accelerated in hardware and/or incorporated into application-specific integrated circuits.

In the future, there may be a host of vision-enabled consumer products, including inexpensive motion capture for video games, autonomous vacuum cleaners and lawn mowers, and safer automobiles. Existing applications in factory automation and surveillance will become much more robust through the expanded use of real-time 3-D technology. □

Meet the author

Ron Buck is CEO of Tyzx Inc. in Menlo Park, Calif.; e-mail: ron@tyzx.com.

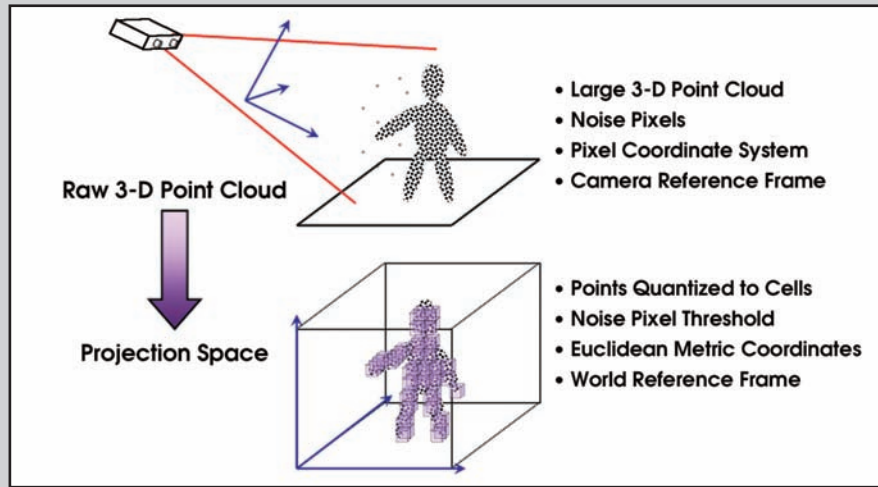


Figure 3. The postprocessor transforms 3-D data into a Euclidean 3-D quantized volume and produces a data representation that is useful for many applications and that is more compact than a full 3-D cloud of points. The primitive provides a mechanism to define and create 3-D projections where clipping and quantization parameters can be defined independently for each of the three axes.